

GRAM SCHMIDT BASED GREEDY HYBRID PRECODING FOR FREQUENCY SELECTIVE MILLIMETER WAVE MIMO SYSTEMS

Ahmed Alkhateeb and Robert W. Heath Jr.

The University of Texas at Austin, TX, USA, Email: {aalkhateeb, rheath}@utexas.edu

ABSTRACT

Hybrid analog/digital precoding allows millimeter wave MIMO systems to leverage large antenna array gains while permitting low cost and power consumption hardware. Most prior work has focused on hybrid precoding for narrow-band mmWave systems. MmWave systems, however, will likely operate on wideband channels with frequency selectivity. Therefore, this paper considers frequency selective hybrid precoding with RF beamforming vectors taken from a quantized codebook. For this system, a low-complexity yet near-optimal greedy algorithm is developed for the design of the hybrid analog/digital precoders. The proposed algorithm greedily selects the RF beamforming vectors using Gram-Schmidt orthogonalization. Simulation results show that the developed precoding design algorithm achieves very good performance compared with the unconstrained solutions while requiring less complexity.

Index Terms— Millimeter wave communications, frequency selective, Gram-Schmidt, hybrid precoding.

1. INTRODUCTION

Millimeter wave (mmWave) communication can leverage the large bandwidth potentially available at the high frequency bands to provide high data rates [1–5]. To guarantee sufficient received signal power at these high frequencies, though, large antenna arrays need to be deployed at both the transmitter and receiver [2, 3, 6]. Designing precoding and combining matrices for these mmWave wideband large MIMO systems differs from lower-frequency solutions. This is mainly due to the different hardware constraints on the mixed signal components because of their high cost and power consumption [7]. Therefore, developing precoding schemes for wideband mmWave systems is important for building these systems.

For the sake of low power consumption, hybrid analog/digital precoding solutions, that divide the precoding between analog and digital domains, and hence requiring smaller number of RF chains, were proposed in [8–16]. For general MIMO systems, hybrid precoding design with diversity and spatial multiplexing objectives were investigated in [8, 9]. In [10], the sparse nature of mmWave channels was exploited, and low-complexity iterative algorithms based on matching pursuit were devised, assuming perfect channel knowledge at the transmitter. Extensions to the case when only partial channel knowledge is required was considered in [11, 12]. Other heuristic algorithms that do not rely on orthogonal matching pursuit were also proposed in [13–15] for the hybrid precoding design with perfect channel knowledge at the transmitter. The work in [10–14] assumed a narrow-band mmWave channel, with perfect or partial

channel knowledge at the transmitter. In [16], hybrid beamforming with only a single-stream transmission over MIMO-OFDM system was considered. The solution in [16] though relied on the joint exhaustive search over both RF and baseband codebooks which results in high-complexity. As mmWave communication is expected to employ broadband channels, developing spatial multiplexing hybrid precoding algorithms for wideband mmWave systems is important.

In this paper, we investigate the frequency selective hybrid precoding design to maximize the achievable mutual information given that the RF precoders are taken from a quantized codebook. We first derive the optimal baseband precoders as functions of the RF precoders. Then, we design a greedy hybrid precoding algorithm based on Gram-Schmidt orthogonalization. Despite its low-complexity, the proposed algorithm is illustrated to achieve a similar performance compared with the optimal hybrid precoding design that requires an exhaustive search over the RF codebooks.

2. SYSTEM MODEL

Consider the OFDM based system model in Fig. 1 where a base station (BS) with N_{BS} antennas and N_{RF} RF chains is assumed to communicate with a single mobile station (MS) with N_{MS} antennas and N_{RF} RF chains. The BS and MS communicate via N_S length- K data symbol blocks, such that $N_S \leq N_{RF} \leq N_{BS}$ and $N_S \leq N_{RF} \leq N_{MS}$.

At the transmitter, the N_S data symbols \mathbf{s}_k at each subcarrier $k = 1, \dots, K$ are first precoded using an $N_{RF} \times N_S$ digital precoding matrix $\mathbf{F}[k]$, and the symbol blocks are transformed to the time-domain using N_{RF} K -point IFFT's. Note that our model assumes that all subcarriers are used and, therefore, the data block length is equal to the number of subcarriers. A cyclic prefix of length D is then added to the symbol blocks before applying the $N_{BS} \times N_{RF}$ RF precoding \mathbf{F}_{RF} . It is important to emphasize here that the RF precoding matrix \mathbf{F}_{RF} is the same for *all* subcarriers. This means that the RF precoder is assumed to be frequency flat while the baseband precoders can be different for each subcarrier. The discrete-time transmitted complex baseband signal at subcarrier k can therefore be written as

$$\mathbf{y}[k] = \mathbf{F}_{RF} \mathbf{F}[k] \mathbf{s}[k], \quad (1)$$

where $\mathbf{s}[k]$ is the $N_S \times 1$ transmitted vector at subcarrier k , such that $\mathbb{E}[\mathbf{s}[k] \mathbf{s}[k]^*] = \frac{P}{K N_S} \mathbf{I}_{N_S}$, and P is the average total transmit power. Since \mathbf{F}_{RF} is implemented using analog phase shifters, its entries are of constant modulus. To reflect that, we normalize the entries $[\mathbf{F}_{RF}]_{m,n} = 1$. Further, we assume that the angles of the analog phase shifters are quantized and have a finite set of possible values. With these assumptions, $[\mathbf{F}_{RF}]_{m,n} = e^{j\phi_{m,n}}$, where $\phi_{m,n}$ is a quantized angle. The hybrid precoders are assumed to have a unitary power constraint, i.e.,

This work is supported in part by the National Science Foundation under Grant No. 1319556, and by a gift from Nokia.

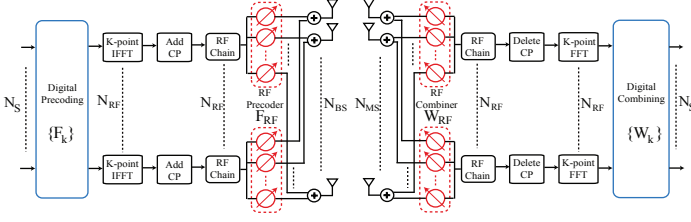


Fig. 1. A block diagram of the OFDM based BS-MS transceiver that employs hybrid analog/digital precoding.

they meet $\mathbf{F}_{\text{RF}}\mathbf{F}[k] \in \mathcal{U}_{N_{\text{BS}} \times N_{\text{S}}}$, with the set of semi-unitary matrices $\mathcal{U}_{N_{\text{BS}} \times N_{\text{S}}} = \{\mathbf{U} \in \mathbb{C}^{N_{\text{BS}} \times N_{\text{S}}} | \mathbf{U}^* \mathbf{U} = \mathbf{I}\}$.

At the MS, assuming perfect carrier and frequency offset synchronization, the received signal is first combined in the RF domain using the $N_{\text{MS}} \times N_{\text{RF}}$ combining matrix \mathbf{W}_{RF} . Then, the cyclic prefix is removed, and the symbols are returned back to the frequency domain where the symbols at each subcarrier k are combined using the $N_{\text{RF}} \times N_{\text{S}}$ digital combining matrix $\mathbf{W}[k]$. Denoting the $N_{\text{MS}} \times N_{\text{BS}}$ channel matrix at subcarrier k as $\mathbf{H}[k]$, the received signal at subcarrier k after processing can be then expressed as

$$\mathbf{y}[k] = \mathbf{W}[k]^* \mathbf{W}_{\text{RF}}^* \mathbf{H}[k] \mathbf{F}_{\text{RF}} \mathbf{F}[k] \mathbf{s}[k] + \mathbf{W}[k]^* \mathbf{W}_{\text{RF}}^* \mathbf{n}[k], \quad (2)$$

where $\mathbf{n}[k] \sim \mathcal{N}(\mathbf{0}, \sigma_{\text{N}}^2 \mathbf{I})$ is a Gaussian noise vector.

3. PROBLEM FORMULATION

The paper objective is to develop a low-complexity hybrid precoding design to maximize the achievable system spectral efficiency. Given the system model in Section 2. For simplicity of exposition, we will assume that the receiver can perform optimal nearest neighbor decoding based on the N_{MS} -dimensional received signal with fully digital hardware. This allows decoupling the transceiver design problem, and focusing on the hybrid precoders design to maximize the mutual information of the system [10], defined as

$$\mathcal{I}(\mathbf{F}_{\text{RF}}, \{\mathbf{F}[k]\}_{k=1}^K) = \frac{1}{K} \sum_{k=1}^K \log_2 \left| \mathbf{I}_{N_{\text{MS}}} + \frac{\rho}{N_{\text{S}}} \mathbf{H}[k] \mathbf{F}_{\text{RF}} \mathbf{F}[k] \mathbf{F}[k]^* \mathbf{F}_{\text{RF}}^* \mathbf{H}[k]^* \right|, \quad (3)$$

where $\rho = \frac{P}{K\sigma^2}$ is the SNR. As combining with fully digital hardware is not a practical mmWave solution, the hybrid combining design problem needs also to be considered. The design ideas that will be given in this paper for the hybrid precoders, however, provide direct tools for constructing the hybrid combining matrices, \mathbf{W}_{RF} , $\{\mathbf{W}[k]\}_{k=1}^K$, and is therefore omitted due to space limitations.

If the RF beamforming vectors are taken from a codebook \mathcal{F}_{RF} that captures the RF hardware constraints, then the maximum mutual information under the given hybrid precoding model is

$$\begin{aligned} \mathcal{I}_{\text{HP}}^* &= \max_{\mathbf{F}_{\text{RF}}, \{\mathbf{F}[k]\}_{k=1}^K} \mathcal{I}(\mathbf{F}_{\text{RF}}, \{\mathbf{F}[k]\}_{k=1}^K) \\ \text{s.t. } &[\mathbf{F}_{\text{RF}}]_{:,r} \in \mathcal{F}_{\text{RF}}, \quad r = 1, \dots, N_{\text{RF}} \\ &\mathbf{F}_{\text{RF}} \mathbf{F}[k] \in \mathcal{U}_{N_{\text{BS}} \times N_{\text{RF}}}, \quad k = 1, 2, \dots, K. \end{aligned} \quad (4)$$

One challenge of the hybrid precoding design to solve the optimization problem in (4) is the coupling between baseband and RF precoders that arises in the power constraint (the second constraint

of (4)). In the following proposition, we show that the baseband precoders can be written optimally as a function of the RF precoders.

Proposition 1 Define the SVD decompositions of the matrices $\mathbf{H}[k] = \mathbf{U}[k] \mathbf{\Sigma}[k] \mathbf{V}[k]^*$ and $\mathbf{\Sigma}[k] \mathbf{V}[k]^* \mathbf{F}_{\text{RF}} (\mathbf{F}_{\text{RF}}^* \mathbf{F}_{\text{RF}})^{-\frac{1}{2}} = \bar{\mathbf{U}}[k] \bar{\mathbf{\Sigma}}[k] \bar{\mathbf{V}}[k]^*$, then the baseband precoders $\{\mathbf{F}[k]\}_{k=1}^K$ that solve (4) are given by

$$\mathbf{F}[k]^* = (\mathbf{F}_{\text{RF}}^* \mathbf{F}_{\text{RF}})^{-\frac{1}{2}} [\bar{\mathbf{V}}[k]]_{:,1:N_{\text{S}}}, \quad k = 1, 2, \dots, K. \quad (5)$$

Proof: The proof follows using change of variables. It is omitted due to space limitation, but available in the journal version [17]. \square

Given proposition 1, the optimal hybrid precoding based mutual information can be given by making an exhaustive search over only the RF precoding codebook. To avoid this search, we propose efficient greedy hybrid precoding algorithms in the following sections.

4. GREEDY HYBRID PRECODING

A natural greedy approach to construct the hybrid precoder is to iteratively select the N_{RF} RF beamforming vectors from the codebook \mathcal{F}_{RF} to maximize the mutual information. In this paper, we call this the direct greedy hybrid precoding (DG-HP) algorithm. Let the $N_{\text{BS}} \times (i-1)$ matrix $\mathbf{F}_{\text{RF}}^{(i-1)}$ denote the RF precoding matrix at the end of the $(i-1)$ th iteration. Then by leveraging the optimal baseband precoder structure in (5), the objective of the i th iteration is to select $\mathbf{f}_{n^*}^{\text{RF}} \in \mathcal{F}_{\text{RF}}$ that solves

$$\begin{aligned} \mathcal{I}_{\text{HP}}^{(i)} &= \max_{\mathbf{f}_{n^*}^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_{\text{RF}}} \right. \\ &\quad \times \lambda_{\ell} \left(\mathbf{H}[k] \hat{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\hat{\mathbf{F}}_{\text{RF}}^{(i,n)*} \hat{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-1} \hat{\mathbf{F}}_{\text{RF}}^{(i,n)*} \mathbf{H}[k]^* \right) \left. \right), \quad (6) \end{aligned}$$

with $\hat{\mathbf{F}}_{\text{RF}}^{(i,n)} = [\mathbf{F}_{\text{RF}}^{(i-1)}, \mathbf{f}_{n^*}^{\text{RF}}]$. The best vector $\mathbf{f}_{n^*}^{\text{RF}}$ will be then added to the RF precoding matrix to form $\mathbf{F}_{\text{RF}}^{(i)} = [\mathbf{F}_{\text{RF}}^{(i-1)}, \mathbf{f}_{n^*}^{\text{RF}}]$. The achievable mutual information with this algorithm is then $\mathcal{I}_{\text{HP}}^{\text{DG-HP}} = \mathcal{I}_{\text{HP}}^{(N_{\text{RF}})}$. The main limitation of this algorithm is that it still requires an exhaustive search over \mathcal{F}_{RF} and eigenvalues calculation in each iteration. In the next section, we will make a first step towards developing a low-complexity algorithm that has a similar (or very close) performance to this DG-HP algorithm.

5. GRAM-SCHMIDT GREEDY HYBRID PRECODING

In hybrid analog/digital precoding architectures, the effective channel seen at the baseband is through the RF precoders lens. This gives the intuition that it is better for the RF beamforming vectors to be orthogonal (or close to orthogonal), as this physically means that the effective channel will have a better coverage over the dominant subspaces belonging to the actual channel matrix. This intuition is also confirmed by the structure of the optimal baseband precoder discussed in Proposition 1, as the overall matrix $\mathbf{F}_{\text{RF}} (\mathbf{F}_{\text{RF}}^* \mathbf{F}_{\text{RF}})^{-\frac{1}{2}}$ has a semi-unitary structure. This note means that in each iteration i of the greedy hybrid precoding algorithm in (6) with a selected codeword $\mathbf{f}_{n^*}^{\text{RF}}$, the additional mutual information gain over the previous iterations is due to the contribution of the component of $\mathbf{f}_{n^*}^{\text{RF}}$ that is orthogonal on the existing RF precoding matrix $\mathbf{F}_{\text{RF}}^{(i-1)}$. This is similar to the greedy user

scheduling in MIMO broadcast channels based on the orthogonal channel components [18], but in a different context. Based on that, we modify the DG-HP algorithm by adding a Gram-Schmidt orthogonalization step in each iteration i to project the candidate beamforming codewords on the orthogonal complement of the subspace spanned by the selected codewords in $\mathbf{F}_{\text{RF}}^{(i-1)}$. This can be simply done by multiplying the candidate vectors by the projection matrix $\mathbf{P}^{(i-1)\perp} = \left(\mathbf{I}_i - \mathbf{F}_{\text{RF}}^{(i-1)} \left(\mathbf{F}_{\text{RF}}^{(i-1)*} \mathbf{F}_{\text{RF}}^{(i-1)} \right)^{-1} \mathbf{F}_{\text{RF}}^{(i-1)*} \right)$. Given the optimal precoder design in (5), the mutual information at the i th iteration of the modified Gram-Schmidt hybrid precoding (GS-HP) algorithm can be written as

$$\begin{aligned} \bar{\mathcal{I}}_{\text{HP}}^{(i)} &= \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_{\text{RF}}} \lambda_{\ell} \left(\mathbf{H}[k] \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right. \right. \\ &\quad \times \left. \left. \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-1} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \mathbf{H}[k]^* \right) \right), \quad (7) \\ &\stackrel{(a)}{=} \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_{\text{RF}}} \lambda_{\ell} \left(\mathbf{T}^{(i-1)} \right. \right. \\ &\quad \times \left. \left. \mathbf{H}[k] \mathbf{P}^{(i-1)\perp} \mathbf{f}_n^{\text{RF}} \mathbf{f}_n^{\text{RF}*} \mathbf{P}^{(i-1)\perp*} \mathbf{H}^*[k] \right) \right), \quad (8) \end{aligned}$$

with $\mathbf{T}^{(i-1)} = \mathbf{H}[k] \mathbf{F}_{\text{RF}}^{(i-1)} \left(\mathbf{F}_{\text{RF}}^{(i-1)*} \mathbf{F}_{\text{RF}}^{(i-1)} \right)^{-1} \mathbf{F}_{\text{RF}}^{(i-1)*} \mathbf{H}[k]^*$, and $\bar{\mathbf{F}}_{\text{RF}}^{(i,n)} = \left[\mathbf{F}_{\text{RF}}^{(i-1)}, \mathbf{P}^{(i-1)\perp} \mathbf{f}_n^{\text{RF}} \right]$. Note that $\mathbf{T}^{(i-1)}$ is a constant matrix at iteration i , and (a) follows from the Gram-Schmidt orthogonalization which allows the matrix $\bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-\frac{1}{2}}$ to be written as $\left[\mathbf{F}_{\text{RF}}^{(i-1)} \left(\mathbf{F}_{\text{RF}}^{(i-1)*} \mathbf{F}_{\text{RF}}^{(i-1)} \right)^{-\frac{1}{2}}, \mathbf{P}^{(i-1)\perp} \mathbf{f}_n^{\text{RF}} \right]$. Hence, the eigenvalues calculation in (8) can be calculated as a rank-1 update of the previous iteration eigenvalues, which reduces the overall complexity [19]. The best vector $\mathbf{f}_{n^*}^{\text{RF}}$ will be then added to the RF precoding matrix to form $\mathbf{F}_{\text{RF}}^{(i)} = \left[\mathbf{F}_{\text{RF}}^{(i-1)}, \mathbf{f}_{n^*}^{\text{RF}} \right]$. At the end of the N_{RF} iterations, we get $\mathcal{I}_{\text{HP}}^{\text{GS-HP}} = \bar{\mathcal{I}}_{\text{HP}}^{(N_{\text{RF}})}$. In the following proposition, we prove that this Gram-Schmidt hybrid precoding algorithm is exactly equivalent to the DG-HP algorithm.

Proposition 2 *The achieved mutual information of the direct greedy hybrid precoding algorithm in (6) and the Gram-Schmidt hybrid precoding algorithm in (7) are exactly equal, i.e., $\mathcal{I}_{\text{HP}}^{\text{DG-HP}} = \mathcal{I}_{\text{HP}}^{\text{GS-HP}}$.*

Proof: See Section 9. \square

6. APPROXIMATE GRAM-SCHMIDT BASED GREEDY HYBRID PRECODING

The main advantage of the Gram-Schmidt hybrid precoding design in Section 5 is that it leads to a near-optimal low-complexity design of the frequency selective hybrid precoding as will be discussed in this section. Given the optimal baseband precoding solution in (5), the mutual information at the i th iteration in (7) can be written as

$$\begin{aligned} \bar{\mathcal{I}}_{\text{HP}}^{(i)} &= \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_{\text{RF}}} \lambda_{\ell} \left(\mathbf{H}[k] \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right. \right. \\ &\quad \times \left. \left. \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-1} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \mathbf{H}[k]^* \right) \right), \quad (9) \end{aligned}$$

Algorithm 1 Approximate Gram-Schmidt Greedy Hybrid Precoding

Initialization

1) Construct $\mathbf{\Pi} = \tilde{\mathbf{\Sigma}}_{\text{H}} \tilde{\mathbf{V}}_{\text{H}}$, with $\tilde{\mathbf{\Sigma}}_{\text{H}} = [\tilde{\mathbf{\Sigma}}_1, \dots, \tilde{\mathbf{\Sigma}}_K]$ and $\tilde{\mathbf{V}}_{\text{H}} = [\tilde{\mathbf{V}}_1, \dots, \tilde{\mathbf{V}}_K]$. Set $\mathbf{F}_{\text{RF}} = \text{Empty Matrix}$. Set $\mathbf{A}_{\text{CB}} = [\mathbf{f}_1^{\text{RF}}, \dots, \mathbf{f}_{N_{\text{CB}}}^{\text{RF}}]$, where $\mathbf{f}_n^{\text{RF}}, n = 1, \dots, N_{\text{CB}}^{\text{v}}$ are the codewords in \mathcal{F}_{RF} .

RF Precoder Design

2) For $i, i = 1, \dots, N_{\text{RF}}$
a) $\mathbf{\Psi} = \mathbf{\Pi}^* \mathbf{A}_{\text{CB}}$
b) $n^* = \arg \max_{n=1,2,\dots,N_{\text{CB}}^{\text{v}}} \left\| [\mathbf{\Psi}]_{:,n} \right\|_2$.
c) $\mathbf{F}_{\text{RF}}^{(i)} = \left[\mathbf{F}_{\text{RF}}^{(i-1)}, \mathbf{f}_{n^*}^{\text{RF}} \right]$
d) $\mathbf{\Pi} = \mathbf{\Pi} \left(\mathbf{I}_i - \mathbf{F}_{\text{RF}}^{(i)} \left(\mathbf{F}_{\text{RF}}^{(i)*} \mathbf{F}_{\text{RF}}^{(i)} \right)^{-1} \mathbf{F}_{\text{RF}}^{(i)*} \right)$

Digital Precoder Design

3) $\mathbf{F}[k] = \mathbf{F}_{\text{RF}}^{(N_{\text{RF}})} \left(\mathbf{F}_{\text{RF}}^{(N_{\text{RF}})*} \mathbf{F}_{\text{RF}}^{(N_{\text{RF}})} \right)^{-\frac{1}{2}} [\bar{\mathbf{V}}[k]]_{:,1:N_{\text{S}}}, k = 1, \dots, K$, with $\bar{\mathbf{V}}[k]$ defined in (5).

$$\begin{aligned} &\stackrel{(a)}{\geq} \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_{\text{S}}} \lambda_{\ell} \left(\tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}^*[k] \right. \right. \\ &\quad \times \left. \left. \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-1} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \tilde{\mathbf{V}}[k] \tilde{\mathbf{\Sigma}}^*[k] \right) \right), \quad (10) \\ &\stackrel{(b)}{\approx} \frac{1}{K} \sum_{k=1}^K \left(\log_2 \left| \mathbf{I} + \frac{\rho}{N_{\text{S}}} \tilde{\mathbf{\Sigma}}[k]^2 \right| - \text{tr} \left(\tilde{\mathbf{\Sigma}}[k] \right) \right) \\ &\quad + \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \left\| \tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}[k]^* \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-\frac{1}{2}} \right\|_F^2 \quad (11) \end{aligned}$$

where (a) is by considering only the first N_{S} dominant singular values of $\mathbf{H}[k]$, $\tilde{\mathbf{\Sigma}}[k] = [\mathbf{\Sigma}[k]]_{:,1:N_{\text{S}}}$, $\tilde{\mathbf{V}}[k] = [\mathbf{V}[k]]_{:,1:N_{\text{S}}}$, and (b) follows from using the large mmWave MIMO approximations used in [10]. The objective of the i th iteration is then to solve

$$\mathbf{f}_{n^*}^{\text{RF}} = \arg \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \frac{1}{K} \sum_{k=1}^K \left\| \tilde{\mathbf{\Sigma}}[k] \tilde{\mathbf{V}}[k]^* \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \left(\bar{\mathbf{F}}_{\text{RF}}^{(i,n)*} \bar{\mathbf{F}}_{\text{RF}}^{(i,n)} \right)^{-\frac{1}{2}} \right\|_F^2 \quad (12)$$

$$= \arg \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \left\| \tilde{\mathbf{\Sigma}}_{\text{H}} \tilde{\mathbf{V}}_{\text{H}}^* \mathbf{F}_{\text{RF}}^{(i-1)} \left(\mathbf{F}_{\text{RF}}^{(i-1)*} \mathbf{F}_{\text{RF}}^{(i-1)} \right)^{-\frac{1}{2}} \right\|_F^2, \quad (13)$$

$$\stackrel{(a)}{=} \arg \max_{\mathbf{f}_n^{\text{RF}} \in \mathcal{F}_{\text{RF}}} \left\| \tilde{\mathbf{\Sigma}}_{\text{H}} \tilde{\mathbf{V}}_{\text{H}}^* \mathbf{P}^{(i-1)\perp} \mathbf{f}_n^{\text{RF}} \right\|_2^2, \quad (14)$$

where $\tilde{\mathbf{\Sigma}}_{\text{H}} = [\tilde{\mathbf{\Sigma}}[1], \dots, \tilde{\mathbf{\Sigma}}[K]]$, $\tilde{\mathbf{V}}_{\text{H}} = [\tilde{\mathbf{V}}[1], \dots, \tilde{\mathbf{V}}[K]]$, and (a) is a result of the Gram-Schmidt processing as described in Section 5. The problem in (14) is simple to solve with just a maximum projection step. We call this algorithm the approximate Gram-Schmidt hybrid precoding (Approximate GS-HP) algorithm. As shown in Algorithm 1, the developed algorithm sequentially build the RF and baseband precoding matrices in two separate stages. First, the RF beamforming vectors are iteratively selected to solve (14). Then, the baseband precoder is optimally designed according to (5). Despite its sequential design of the RF and baseband precoders, which reduces the complexity when compared with prior solutions that mostly depend on the joint design of the baseband and RF precoding matrices [10, 12], Algorithm 1 achieves a significant gain over prior

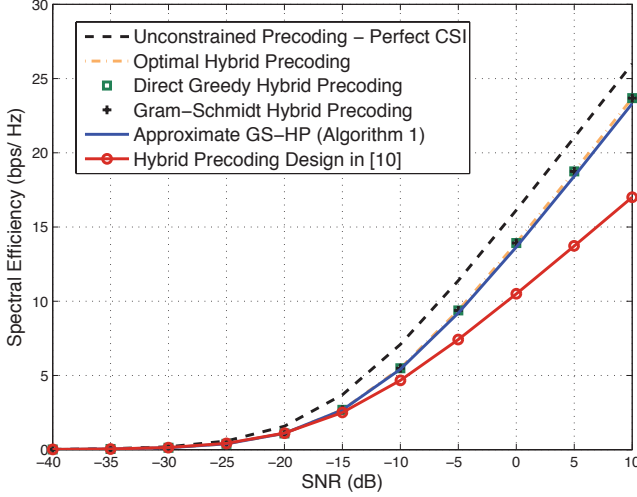


Fig. 2. The performance of the approximate Gram-Schmidt hybrid precoding design in Algorithm 1 compared with the optimal hybrid precoding solution, the unconstrained SVD solution, and the prior work in [10]. The system has $N_{BS} = 32$ antennas, $N_{MS} = 16$ antennas, and $N_S = N_{RF} = 3$.

solutions, and gives a very close performance to the optimal solution given by exhaustive search, as will be shown in Section 7.

7. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed algorithm using numerical simulations. We adopt a wideband mmWave channel model that consists of $L = 6$ clusters. The center AoAs/AoDs of the L clusters θ_ℓ, ϕ_ℓ are assumed to be uniformly distributed in $[0, 2\pi)$. Each cluster has $R_\ell = 5$ rays with Laplacian distributed AoAs/AoDs [10, 20], and angle spread of 10° . The number of system subcarriers K equals 512, and the cyclic prefix length is $D = 128$, which is similar to 802.11ad [21]. The paths delay is uniformly distributed in $[0, DT_s]$. Both the BS and MS have 12 antennas. In addition to evaluating the approximate Gram-Schmidt based hybrid precoding algorithm. The spectral efficiencies achieved by these greedy algorithms are compared with the optimal hybrid precoding design given by the exhaustive search over the RF codebooks. The rates are also compared with the prior solution in [10]. For a fair comparison, we assume that each RF beamforming vector is selected from a beamsteering codebook with a size $N_{CB} = 64$. First, Fig. 2 shows that the direct greedy and Gram-Schmidt based hybrid precoding algorithms achieve exactly the same performance which verifies Proposition 2. Despite its low-complexity, the developed approximate Gram-Schmidt hybrid precoding design in Algorithm 1 achieves very close performance to the exhaustive-search based optimal solution. We emphasize here that any hybrid precoding design can not perform better than the shown optimal hybrid precoding solution with the considered RF codebook, which confirms the near-optimal result of the proposed algorithm. This is also clear from the considerable gain obtained by the proposed algorithm compared with the prior solution in [10]. Also, it is worth mentioning that the developed hybrid precoding algorithms in this paper can be applied to any large MIMO system (not specifically mmWave systems).

8. CONCLUSION

In this paper, we investigated hybrid precoding design for wideband mmWave systems. First, we derived the optimal hybrid precoding design that maximizes the achievable mutual information for any given RF codebook, and showed that the optimal baseband structure can be decomposed into an RF precoder dependent matrix and a unitary matrix. Second, we developed a novel greedy hybrid precoding algorithm based on Gram-Schmidt orthogonalization. Thanks to this Gram-Schmidt orthogonalization, we showed that only sequential design of the RF and baseband precoders is required to achieve the same performance of more sophisticated algorithms that requires a joint design of the RF and baseband precoders in each step. Simulation results illustrated that the proposed precoding algorithms improve over prior work and stay within a small gap from the unconstrained perfect channel knowledge solutions.

9. PROOF OF PROPOSITION 2

Proof: To prove that $\mathcal{I}_{HP}^{GS-HP} = \mathcal{I}_{HP}^{DG-HP}$, it is sufficient to prove that $\mathbf{F}_{RF}^{(N_{RF})}$ of the GS-HP and DG-HP algorithms are equal. To do that, we will show that both the algorithms choose the same RF beamforming vector in each iteration, i.e., $\mathbf{F}_{RF}^{(i)}$ is equal for $i = 1, \dots, N_{RF}$. This can be proved using mathematical induction as follows. At the first iteration, the two algorithms do the exhaustive search over the same codebook \mathcal{F}_{RF} , and consequently select the same beamforming vectors. Now, suppose that the two algorithms reach the same RF precoding matrix $\mathbf{F}_{RF}^{(i-1)}$ at iteration $i - 1$, we need to prove that they both select the same RF beamforming vector at iteration i , i.e., we need to prove that both (6) and (7) choose beamforming vectors with the same index. To prove that, it is enough to show that the contributions of the n th beamforming vector \mathbf{f}_n^{RF} from \mathcal{F}_{RF} in (6) and (7) are equal. Given the optimal baseband precoder in (5), and denoting the SVD of $\hat{\mathbf{F}}_{RF}^{(i,n)}$ as $\hat{\mathbf{F}}_{RF}^{(i,n)} = \hat{\mathbf{U}}_{RF}^{(i,n)} \hat{\Sigma}_{RF}^{(i,n)} \hat{\mathbf{V}}_{RF}^{(i,n)*}$, equation (6) can be written as

$$\sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_S} \lambda_\ell \left(\mathbf{H}[k] \hat{\mathbf{F}}_{RF}^{(i,n)} \left(\hat{\mathbf{F}}_{RF}^{(i,n)*} \hat{\mathbf{F}}_{RF}^{(i,n)} \right)^{-1} \times \hat{\mathbf{F}}_{RF}^{(i,n)*} \mathbf{H}[k]^* \right) \right), \quad (15)$$

$$= \sum_{\ell=1}^i \log_2 \left(1 + \frac{\rho}{N_S} \lambda_\ell \left(\mathbf{H}[k] \hat{\mathbf{U}}_{RF}^{(i,n)} \hat{\mathbf{U}}_{RF}^{(i,n)*} \mathbf{H}[k]^* \right) \right). \quad (16)$$

Equation (7) can be similarly written, but with $\hat{\mathbf{U}}_{RF}^{(i,n)}$ replaced by $\bar{\mathbf{U}}_{RF}^{(i,n)}$ where $\bar{\mathbf{F}}_{RF}^{(i,n)} = \bar{\mathbf{U}}_{RF}^{(i,n)} \bar{\Sigma}_{RF}^{(i,n)} \bar{\mathbf{V}}_{RF}^{(i,n)*}$. Hence, we need to prove that $\hat{\mathbf{U}}_{RF}^{(i,n)} \hat{\mathbf{U}}_{RF}^{(i,n)*} = \bar{\mathbf{U}}_{RF}^{(i,n)} \bar{\mathbf{U}}_{RF}^{(i,n)*}$. Let $\bar{\mathbf{f}}_n = \mathbf{P}^{(i-1)\perp} \mathbf{f}_n$ denote the last column of $\bar{\mathbf{F}}_{RF}^{(i,n)}$. As $\bar{\mathbf{f}}_n$ is a result of successive Gram-Schmidt operations, we can write $\mathbf{f}_n = \bar{\mathbf{f}}_n + \mathbf{F}_{RF}^{(i-1)} \alpha_n$, where α_n is a vector results from the Gram-Schmidt process. Consequently, $\hat{\mathbf{F}}_{RF}^{(i,n)}$ can be written as $\hat{\mathbf{F}}_{RF}^{(i,n)} = \bar{\mathbf{F}}_{RF}^{(i,n)} \mathbf{E}_C$, where $\mathbf{E}_C = \begin{bmatrix} \mathbf{I} & \alpha \\ \mathbf{0}^T & 1 \end{bmatrix}$ is an elementary column operation matrix. Now, we note that $\hat{\mathbf{U}}_{RF}^{(i,n)} \hat{\mathbf{U}}_{RF}^{(i,n)*} = \hat{\mathbf{F}}_{RF}^{(i,n)} \hat{\mathbf{F}}_{RF}^{(i,n)\dagger} = \bar{\mathbf{F}}_{RF}^{(i,n)} \mathbf{E}_C \mathbf{E}_C^{-1} \bar{\mathbf{F}}_{RF}^{(i,n)\dagger} = \bar{\mathbf{U}}_{RF}^{(i,n)} \bar{\mathbf{U}}_{RF}^{(i,n)*}$, as \mathbf{E}_C is an $i \times i$ full-rank matrix. \square

10. REFERENCES

- [1] T. Nitsche, C. Cordeiro, A. Flores, E. Knightly, E. Perahia, and J. Widmer, "IEEE 802.11ad: directional 60 GHz communication for multi-gigabit-per-second Wi-Fi," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, December 2014.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.
- [3] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Pearson Education, 2014.
- [4] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [5] T. Bai, A. Alkhateeb, and R. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 70–77, Sept. 2014.
- [6] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [7] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, December 2014.
- [8] J. Zhang, Y. Wu, S. Zhou, and J. Wang, "Joint linear transmitter and receiver design for the downlink of multiuser MIMO systems," *IEEE Communications Letters*, vol. 9, no. 11, pp. 991–993, Nov 2005.
- [9] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.
- [10] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [11] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. of Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–5.
- [12] —, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [13] F. Sahrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *Proc. of the IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015.
- [14] C. Mendez-Rial, R. Rusu, N. Gonzalez-Prelcic, and R. W. Heath, "Dictionary-free hybrid precoders and combiners for mmwave MIMO systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015.
- [15] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Communications Letters*, vol. PP, no. 99, pp. 1–1, 2015.
- [16] C. Kim, T. Kim, and J.-Y. Seol, "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proc. of IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, Dec. 2013, pp. 61–65.
- [17] A. Alkhateeb and R. W. Heath Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *submitted to IEEE Transactions on Communications*, *arXiv preprint arXiv:1510.00609*, Oct. 2015.
- [18] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
- [19] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 20 – 30, 2006, special Issue on Large Scale Linear and Nonlinear Eigenvalue Problems.
- [20] A. Forenza, D. Love, and R. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, pp. 1924–1934, July 2007.
- [21] 11ad, "IEEE 802.11ad standard draft D0.1." [Online]. Available: www.ieee802.org/11/Reports/tgadupdate.htm